

La méthode des Prix Hédoniques

Principes et Illustration à partir du Prix des Terrains à Bâtir *

Ronan Le Saout (INSEE-CREST),

Benjamin Vignolles (Soes et CREST)

Résumé

Nous menons une analyse hédonique des prix des terrains à bâtir sur la période 2006-2012 à l'aide de près de 400.000 observations individuelles. La méthode des prix hédoniques est utilisée à la fois pour mesurer le prix d'aménités environnementales et construire un indice de prix. Les prix des terrains s'expliquent par des caractéristiques localisées, avec un gradient de prix prononcé pour la distance aux aménités ou aux équipements. Si ces aménités sont valorisées dans le prix, elles n'ont néanmoins que peu d'influence sur l'indice des prix. Les indices construits, quelle que soit la méthode retenue, présentent des évolutions différentes de celles des prix moyens. Plusieurs conclusions peuvent être formulées sur la construction d'indices hédoniques. Pour construire un indice hédonique, seule une fonction de prix doit être estimée. Cette fonction permet de définir un prix implicite aux caractéristiques du bien mais ne renseigne pas sur l'offre ou la demande. Les variables relatives aux acheteurs ou aux vendeurs ne doivent donc pas être incluses dans cette fonction de prix. Quatre grandes familles de modèles hédoniques sont identifiées, dont les hypothèses relatives au panier de biens de référence diffèrent. Les indices construits sont sensibles au choix de modèle et au biais de variables omises, i.e. une caractéristique du bien qui est observée par l'acheteur et le vendeur mais pas dans la base de données.

Mots Clés : prix hédoniques, indice de prix, marché de l'immobilier.

Codes JEL Codes : R32, Q51, D01.

*Nous remercions Thomas Balcone, Pauline Givord et Anne Laferrère pour leurs commentaires sur de premières versions de cette communication. Cette étude ne reflète pas les opinions de l'INSEE et du Soes.

1 Introduction

La méthode des prix hédoniques est d'un emploi courant en économie et statistique. L'intuition, simple, est que le prix d'un bien (appartement, voiture, ordinateur...) ou d'un service (salaire) est fonction de ses caractéristiques. Par exemple, le prix d'un logement s'expliquerait par sa surface, son nombre de pièces mais également par les infrastructures ou la pollution présentes dans son voisinage. Cette méthode sert notamment à calculer et corriger des indices de prix ou à définir un prix à des biens non marchands ou des aménités environnementales.

La méthode est simple en première approche. Elle consiste en effet à expliquer les prix d'un bien par ses caractéristiques (et non celles des acheteurs ou des vendeurs), souvent à l'aide des moindres carrés ordinaires. Comme pour toutes applications empiriques, il est nécessaire de réfléchir à la spécification et aux questions traditionnelles de l'économétrie (endogénéité et corrélation des variables explicatives, hétérogénéité des comportements). Mais en tant que telle cette analyse ne nous renseigne en rien sur l'offre et la demande des biens étudiés. Pour calculer un consentement à payer ou les paramètres d'une fonction de demande, d'autres variables (sur les acheteurs) ou hypothèses (sur les fonctions de demande) doivent être utilisées. Les choix à opérer ne seront donc pas les mêmes selon qu'on conduit une étude économique ou qu'on cherche à calculer un indice des prix.

L'objectif de notre étude est de dégager de bonnes pratiques pour la construction d'un indice hédonique et d'identifier les sources de biais tant sur le choix des variables ou l'estimation de la fonction de prix. Le rappel de la théorie économique nous permet de cadrer le choix des variables explicatives utiles pour calculer une fonction de prix hédonique. Elle ne doit pas être expliquée par des variables relatives à la demande ou l'offre. Seules les caractéristiques du bien doivent être prises en compte. Un prix implicite peut ainsi être associé à chaque caractéristique, dont on pourra ensuite dans une seconde étape chercher à expliquer la demande ou l'offre. Pour la construction d'un indice, seule la fonction de prix est estimée. Le choix des variables peut néanmoins demeurer délicat car certaines caractéristiques importantes peuvent ne pas être connues. Nous détaillons une classification des modèles hédoniques pour la construction d'un indice, avec 4 familles identifiées. Ces types de modèles ne reposent pas sur les mêmes hypothèses relatives au panier de biens de référence. Nous illustrons ces points à l'aide de l'enquête sur le prix des terrains à bâtir (EPTB) sur la période 2006-2012. Nous estimons une fonction de prix hédonique avec des caractéristiques de localisation telles que la présence (ou l'accessibilité) d'aménités et d'équipements. Nous construisons les in-

dices des prix des terrains à bâtir afférents. L'indice du prix des terrains à bâtir est sensible à l'omission d'une variable approchant le coefficient d'occupation des sols, qui régleme la surface pouvant être construite sur un terrain. Ce coefficient est en effet défini par chaque commune dans son Plan Local d'Urbanisme (PLU)¹. Par contre, l'indice est peu sensible à l'inclusion de l'ensemble des aménités (présence du haut-débit, lieu touristique) ou d'indicateurs d'accessibilité des équipements (gare, aéroport, hôpital, supermarché). Ces variables expliquent le prix et la valorisation économique des biens mais ne modifient que marginalement l'indice. Les méthodes par variables indicatrices, par définition d'un bien de référence ou à qualité fixée donnent des résultats proches. Par contre, les résultats sont différents avec la méthode par imputation.

De très nombreuses applications de la méthode hédonique existent. L'évaluation des biens non marchands (la vue sur mer, la pollution...) peut être effectuée par différentes méthodes, dont la méthode des prix hédoniques. Plusieurs articles ont traité de zones géographiques françaises à partir du marché de l'immobilier (Bureau et Glachant 2010, Choumet et Travers 2010, Fack et Grenet 2010, Grislain-Letrémy et Katosky 2013 pour des exemples). L'estimation de fonction de demande ou des consentements à payer sont plus rares (Cavailhès 2005, Baudry et al. 2009 pour des exemples), de même que les analyses sur les services.

Moulton (2001) détaille l'usage des méthodes hédoniques pour les indices de prix aux Etats-Unis. Il distingue plusieurs types d'utilisation des méthodes. Tout d'abord elles permettent de construire des indices de prix pour des biens hétérogènes, tels que les logements. La simple comparaison des prix moyens d'une période à l'autre n'aurait pas de sens car la structure des ventes n'a pas de raison d'être homogène entre 2 périodes. L'idée est alors de contrôler les prix des caractéristiques observables pour en déduire une évolution des prix à caractéristiques constantes. De plus, elle permet de prendre en compte les différences de qualité des biens, à travers l'âge ou d'état général d'une habitation par exemple, mais aussi le progrès technique, en décomposant par exemple le prix d'un ordinateur par ces caractéristiques. Enfin, elle sert également à effectuer des ajustements de qualité ou de mode lorsque des produits sont substitués dans un panier de biens. En France, cette méthodologie est notamment utilisée pour construire les indices de prix des logements anciens (INSEE-Notaires, Gouriéroux et Laferrère 2009) et neufs (INSEE, Balcone 2013), des terres et prés agricoles (Lefebvre et Rouquette 2012) et de la construction. Des ajustements de qualité "hédoniques" sont également effectués dans l'indice des prix à la consommation (notamment pour les lave-linges, les réfrigérateurs, et les livres).

1. Le COS a été supprimé par la loi ALUR 2014.

La suite du document est organisée comme suit. La partie 2 décrit les fondements théoriques de la méthode. Elle distingue la conduite d'études économiques de la construction d'indices. Les hypothèses sous-jacentes (parfois non testables) sont mises en avant pour éclairer l'interprétation des résultats. La partie 3 illustre ces fondements théoriques, à travers l'estimation d'une équation hédonique et d'un indice des prix des terrains constructibles. La partie 4 conclut.

2 Du prix d'un bien au prix d'une caractéristique : les fondements théoriques

Le prix d'un bien (appartement, voiture, ordinateur...) ou d'un service (salaire) est une relation d'équilibre entre une offre et une demande. On observe des prix pour différents biens vendus. Pourquoi ces prix sont-ils différents ? Parce qu'ils n'ont pas les mêmes caractéristiques. L'utilité des consommateurs dépendrait en effet de la quantité de chaque caractéristique présente dans le bien. C'est l'hypothèse de base des modèles hédoniques. Elle vise à relier le prix global au prix de chaque caractéristique et à l'utilité afférente pour le consommateur. Elle permet ainsi de calculer le prix implicite de chaque caractéristique, d'évaluer le consentement à payer pour une caractéristique ou de calculer un indice de prix. La méthode des prix hédoniques a fait l'objet d'une vaste littérature depuis les années 1970. Pour des détails plus théoriques, on pourra notamment se référer à Sheppard (1999) et Palmquist (2005).

2.1 L'équilibre offre-demande pour un bien composite

La théorie économique est issue de Lancaster (1966) et Rosen (1974). Un individu i peut consommer un bien composite \underline{z} de caractéristiques z_1, \dots, z_n (par exemple un ordinateur) de prix $P(\underline{z})$ et un bien x représentant tous les autres biens de prix unitaire. Pour choisir son panier de consommation, un individu (dont les caractéristiques et préférences individuelles sont notées α_i) maximise alors son utilité sous une contrainte de revenu :

$$\begin{aligned} \text{Max} \quad & U_i(x, z_1, \dots, z_n; \alpha_i) \\ \text{s.c.} \quad & R_i = x + P(\underline{z}) \end{aligned}$$

Les agents sont preneurs de prix et ne consomment qu'un seul bien \underline{z} . Le taux marginal de substitution entre une caractéristique z_j et x est alors égale au prix implicite de z_j (le prix

de x étant unitaire) $p_j = \frac{\partial P}{\partial z_j} = \frac{\partial U / \partial z_j}{\partial U / \partial x}$.

La fonction de prix est la même pour tous les acheteurs et vendeurs. Elle est non linéaire (i.e. $\frac{\partial P}{\partial z_j}$ varie en fonction du prix global), ce qui permet d'assurer un prix implicite p_j non constant et donc un équilibre qui ne se résume pas à un point unique. Selon les caractéristiques individuelles α_i , l'enchère (ou le montant) $\theta_i(z, u, R_i)$ que sera prêt à payer un individu pour un ensemble de caractéristiques z à un niveau d'utilité u et un revenu R ne sera pas le même. Les quantités et utilités d'équilibre seront tels que $p(z^*)$ et $\theta_i(z^*, u^*, R_i)$ soient tangents.

De manière équivalente, on peut définir une fonction d'offre $\phi_k(z, \pi, \beta_k)$ avec π le profit et β_k les caractéristiques d'une firme k (paramètres de la fonction de production et prix des facteurs, expliquant les coûts de production). Les quantités et profits d'équilibre seront tels que $p(z^*)$ et $\phi_k(z^*, \pi^*, \beta_k)$ soient tangents.

La fonction de prix $p(z)$ peut être analysée comme l'enveloppe des fonctions d'enchère ou d'offre (Baudray et al. 2009 pour une représentation graphique). En tout point, elle représente un équilibre offre-demande mais ne contient pas en tant que tel d'informations sur la demande ou l'offre.

Ces premiers éléments théoriques ont des implications fortes. Le premier point est que la fonction de prix $p(z)$ doit être non linéaire et ne doit pas contenir de variables sur les caractéristiques des acheteurs (α_i) ou des vendeurs (β_k). Expliquer les prix de l'immobilier ou d'un terrain constructible par le revenu ou la catégorie socio-professionnelle des acheteurs ou des vendeurs n'a donc pas de sens économique.

2.2 De la fonction de prix hédonique à l'analyse hédonique des marchés

L'analyse hédonique du marché en termes d'offre et de demande n'est réalisée que dans une seconde étape, à l'aide des prix individuels estimés. Il est en effet possible de régresser les quantités de chaque caractéristique incorporée dans chaque produit vendu sur les prix implicites estimés dans l'étape précédente. Il faut toutefois prendre en compte des problèmes de simultanéité qui peuvent biaiser les estimateurs ainsi obtenus.

Cette simultanéité peut prendre une forme classique d'endogénéité, fréquemment rencontrée dans les problèmes d'estimation de fonctions d'offre et de demande. En effet, seuls les points d'égalisation entre ces deux courbes sont observés. L'offre et la demande s'expliquent chacune par les quantités et les prix. Une simple régression des quantités de caractéristiques

consommées sur leurs prix estimés conduit à des estimateurs biaisés. Il faut dès lors disposer de variables dont les variations n'affectent que l'une des deux fonctions et les introduire dans les régressions spécifiques auxquelles elles se rapportent. Cela se ramène à un problème de variables instrumentales. Offre et demande sont alors identifiées séparément l'une de l'autre grâce au choix d'instruments spécifiques qui n'affectent que l'une ou l'autre séparément. Ces variables peuvent par exemple être les caractéristiques économiques et sociodémographiques des acheteurs et des vendeurs. Il est également possible de ne considérer que l'une des deux fonctions. C'est par exemple le cas pour les modèles de demande de logement, l'offre étant supposée inélastique et exclue du modèle. Ce choix ne fait pas pour autant disparaître le risque d'endogénéité. Un autre cas d'endogénéité est l'erreur de mesure associée au prix implicite, résultat d'une première estimation. La fonction de prix n'est en pratique que rarement estimée pour des valeurs extrêmes des différentes caractéristiques. L'erreur de mesure est alors positivement corrélée au niveau du prix de la caractéristique. La considération de multiples marchés lors de l'estimation de première étape permet d'estimer plus précisément la fonction de prix hédonique et, ce faisant, de réduire cette source de biais.

Il existe des façons alternatives de prendre en compte les effets de demande dans l'estimation des modèles hédoniques. Par exemple, il est possible de recourir à un modèle en deux étapes à effet de sélection de type "Tobit II" (Palmquist 2006). La seconde étape se présente comme une régression hédonique classique. La première étape modélise le choix, discret, d'acheter le produit observé et non un autre. Pour ce faire, il faut dupliquer les observations des caractéristiques des acheteurs et leur associer des produits qu'ils n'ont pas achetés, ce qui permet d'estimer la propension à demander des produits combinant des caractéristiques spécifiques en fonction des caractéristiques des acheteurs.

2.3 De la fonction de prix hédonique au calcul d'un indice de prix hédonique

Par rapport à la théorie économique, la méthode des prix hédoniques n'est pas que la simple application de la régression de première étape, dont l'estimation relève de choix statistiques. Elle fait en effet intervenir une hypothèse sur la composition du panier de biens de référence, hypothèse non prise en compte dans les analyses précédentes. Plusieurs problèmes vont être ici détaillés : les différents types d'indices, le choix des caractéristiques et l'endogénéité, le biais de variables omises, et la prise en compte de l'espace. Les indices de prix hédoniques visent à construire des indices de prix pour des biens composites, mais également à effectuer un ajustement qualité lorsque des biens disparaissent d'un panier de biens

et doivent être remplacés par des biens proches mais non similaires. Nous concluons par une analyse spécifique de l’ajustement qualité.

Pour le calcul d’indice de prix, la majorité des applications s’appuie sur une modélisation linéaire et logarithmique du prix sous la forme $\text{Log}(\text{Prix}) = \beta \cdot X + \varepsilon$ avec X les caractéristiques du bien et ε un terme d’erreur supposé exogène. Cette forme fonctionnelle permet une interprétation économique directe, les paramètres β représentent des élasticités ou des semi-élasticités (une telle interprétation n’est néanmoins pas toujours souhaitable, par exemple pour un bien immobilier la présence d’un garage ne représente pas forcément une part constante du prix). Elle permet de plus d’obtenir un calcul simplifié des indices par période adjacente (cf. infra). Nous détaillerons donc la théorie à partir de cette forme fonctionnelle.

D’autres formes plus flexibles, semi et non-paramétriques, sont bien sûr possibles et pourront être abordées dans l’application. Dans les applications économiques, on considère par exemple des transformations Box-Cox du prix et des variables explicatives continues, sous la forme $(x^\lambda - 1) / \lambda$ si $\lambda \neq 0$ et $\text{Log}(x)$ si $\lambda = 0$. Le cas d’une modélisation logarithmique du prix est alors souvent rejeté (Grislain-Letrémy et Katosky 2013 et Cropper *et al.* 1988). Il y a un gain sur l’ajustement de la fonction hédonique mais pas forcément sur le calcul des prix implicites. Le calcul d’un indice hédonique demeure complexe dans ce cadre et n’est que peu abordé dans la littérature. Une autre solution est d’adopter une forme fonctionnelle semi ou non paramétrique sur les seules variables explicatives, par exemple à l’aide de régressions splines et fragmentées (Dachary-Bernard *et al.* 2011 pour un exemple).

Les éléments théoriques se concentrent sur les indices élémentaires non pondérés. Ces indices peuvent être calculés par strate (géographique ou de produits) et ensuite être agrégés à l’aide d’une pondération adéquate (nombre de transactions ou montant des ventes par exemple). Mais les indices élémentaires peuvent également être pondérés, notamment dans le cas d’utilisation de données d’enquête (pour être représentatif de l’ensemble des ventes) ou d’adéquation aux stocks des logements par exemple (les ventes représentant un flux).

2.3.1 Classification des indices hédoniques

A partir d’une revue de littérature sur les indices des prix des logements dans le monde et de Triplett (2004), Hill (2013) définit une classification des indices de prix hédoniques et leurs propriétés. Quatre grandes familles d’indices hédoniques existent : l’agrégation de périodes adjacentes, l’imputation de prix, la définition d’un bien de référence et l’ajustement à qualité

fixée. Le tableau 1 résume les propriétés de ces 4 méthodes, en liaison avec la théorie des indices et le choix du panier de référence.

Notons au préalable que pour un modèle logarithmique du type $\text{Log}(\text{Prix}) = \beta \cdot X + \varepsilon$ et dans un cadre homoscédastique (i.e. $\mathbb{V}(\varepsilon/X) = \sigma^2$), la prévision optimale $\widehat{\text{Prix}}$ n'est pas $\text{Exp}(\widehat{\beta} \cdot X)$ mais $\text{Exp}(\widehat{\beta} \cdot X) \cdot \text{Exp}(\widehat{\sigma}^2/2)$. En effet, $\mathbb{E}(\text{Log}(\text{Prix})/X) = \text{Log}(\text{Prix})$ mais $\mathbb{E}(\text{Prix}/X) = (\text{Prix}) \cdot \text{Exp}(\sigma^2/2)$ sous hypothèse de normalité et d'homoscédasticité des termes d'erreur. Nous nous interrogerons donc sur la pertinence d'inclure ou non un tel facteur correctif pour le calcul des indices.

La première méthode, la plus simple, consiste à regrouper plusieurs périodes et à estimer un unique modèle hédonique auquel on ajoute des indicatrices temporelles. Lorsque l'ensemble des périodes sont prises en compte, le modèle est $\text{Log}(\text{Prix}) = \beta \cdot X + \delta \cdot D + \varepsilon$ avec X les caractéristiques et D les indicatrices temporelles (avec 0 comme période de référence). On obtient alors :

$$\begin{aligned} I_{0,t}^{\text{PA}} &= \widehat{\text{Prix}}^{(t)} / \widehat{\text{Prix}}^{(0)} \\ &= \left[\text{Exp}(\widehat{\beta} \cdot X + \widehat{\delta}_t) \cdot \text{Exp}(\widehat{\sigma}^2/2) \right] / \left[\text{Exp}(\widehat{\beta} \cdot X) \cdot \text{Exp}(\widehat{\sigma}^2/2) \right] \\ &= \text{Exp}(\widehat{\delta}_t) \end{aligned}$$

On peut de plus montrer que $\mathbb{E} \left[\text{Exp}(\widehat{\delta}_t) \right] = \text{Exp} \left(\delta_t + \frac{1}{2} \mathbb{V}(\widehat{\delta}_t) \right)$. En pratique, ce facteur correctif $\text{Exp} \left(\frac{1}{2} \mathbb{V}(\widehat{\delta}_t) \right)$ est très faible et peut être négligé. La force de la méthode est d'obtenir un indice très simple, qui ne dépend pas directement du panier de référence des biens. La variance de l'indice se déduit alors aisément par Delta-Méthode, $\mathbb{V}(I_{0,t}^{\text{PA}}) = \mathbb{V}(\widehat{\delta}_t) \text{Exp}[2\delta_t]$. Une variante, chaînée, est de calculer cet indice sur 2 périodes adjacentes. Cela permet de relâcher l'hypothèse de valeurs identiques $\widehat{\beta}$ des caractéristiques quelle que soit la période. Cela évite également d'obtenir des indices variant après ajout d'une nouvelle période. En France, c'est le choix retenu pour les indices des terres et prés agricoles (Lefebvre et Rouquette 2012) et des maisons neuves (Balcone 2013). Le lien avec la théorie générale des indices n'est par contre pas immédiat. En notant $I_{0,t}^{\text{PA}} = \frac{p(X^0 \cup t, \beta^t)}{p(X^0 \cup t, \beta^0)} = \frac{p(X^0 \cup t, (\beta_0^t, \beta_{-0}^t))}{p(X^0 \cup t, (\beta_0^0, \beta_{-0}^0))}$ avec p la fonction hédonique, on remarque que l'indice peut s'interpréter comme l'évolution du prix d'un bien moyen sur les périodes de référence (0) et courante (t). Les autres méthodes permettent d'intuiter que cela se rapproche d'un indice de Fisher ou de Törnqvist.

Les autres méthodes s'appuient toutes sur l'estimation d'un modèle hédonique $\text{Log}(\text{Prix})^{(i)} = \beta^{(i)} \cdot X^{(i)} + \varepsilon^{(i)}$ à chaque période i . Ces méthodes sont donc plus coûteuses en termes de données. Il est alors possible de définir les principaux indices usuels, arithmétique ou géométrique.

Un indice de Laspeyres fixe un panier de biens de référence à la période 0 de caractéristiques X^0 , i.e. $\frac{p(X^0, \beta^t)}{p(X^0, \beta^0)}$. Un indice de Paasche fixe au contraire le panier de biens de référence à la période t , i.e. $\frac{p(X^t, \beta^t)}{p(X^t, \beta^0)}$. Un indice de Fisher est la racine carrée des indices arithmétiques de Laspeyres et Paasche. Un indice de Törnqvist est la racine carrée des indices géométriques de Laspeyres et Paasche.

Chacune de ces méthodes revenant à un ratio d'une fonction de prix hédonique à la période t et 0 (resp. t et $t - 1$ pour les indices chaînés), un facteur correctif $\text{Exp} \left(\left[\widehat{\sigma}_t^2 - \widehat{\sigma}_0^2 \right] / 2 \right)$ (resp. $\text{Exp} \left(\left[\widehat{\sigma}_t^2 - \widehat{\sigma}_{t-1}^2 \right] / 2 \right)$) devrait être théoriquement ajouté. Pour simplifier les notations (et au vu des conclusions de la partie appliquée), nous n'en tenons pas compte. La méthode par imputation consiste à estimer pour chaque individu i de la période de référence la différence de prix hédoniques constatés à la période 0 et t . Les indices de Laspeyres et Paasche arithmétique sont alors le ratio de la somme des prix, i.e. $I_{0,t}^{\text{Imputation,L}} = \sum_{i \subset 0} \text{Exp} \left(\left(\widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot X^{(i)} \right)$ et $I_{0,t}^{\text{Imputation,P}} = \sum_{i \subset t} \text{Exp} \left(\left(\widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot X^{(i)} \right)$. La méthode par bien de référence est équivalente à une méthode d'imputation appliquée à des indices géométriques. Cela revient à calculer des biens moyens de référence $\overline{X}^{(0)}$ et $\overline{X}^{(t)}$. Les indices de Laspeyres et Paasche sont alors, i.e. $I_{0,t}^{\text{Bien Ref.,L}} = \text{Exp} \left(\left(\widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot \overline{X}^{(0)} \right)$ et $I_{0,t}^{\text{Bien Ref.,P}} = \text{Exp} \left(\left(\widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot \overline{X}^{(t)} \right)$. La méthode par ajustement à qualité fixée revient à comparer un prix de référence $\alpha^{(i)}$, i.e. la constante ajustée du modèle hédonique $\text{Log}(\text{Prix})^{(i)} = \alpha^{(i)} + \beta^{-(i)} \cdot X^{-(i)} + \varepsilon^{(i)}$, entre les périodes t et 0. Les indices de Laspeyres et Paasche sont alors, i.e. $I_{0,t}^{\text{Qualité Fixée.,L}} = \text{Exp} \left(\left(\overline{\text{Log}(\text{Prix})}^{(t)} - \widehat{\beta}^{-(0)} \overline{X}^{-(t)} \right) - \widehat{\alpha}^{(0)} \right)$ et $I_{0,t}^{\text{Qualité Fixée.,P}} = \text{Exp} \left(\widehat{\alpha}^{(t)} - \left(\overline{\text{Log}(\text{Prix})}^{(0)} - \widehat{\beta}^{-(t)} \overline{X}^{-(0)} \right) \right)$. Les indices par définition d'un bien de référence ou par ajustement à qualité fixée peuvent s'interpréter comme des indices géométriques. Hill (2011) les interprète au contraire comme des indices arithmétiques sur des biens moyens. Le tableau 1 conserve cette convention. Les méthodes par définition d'un bien de référence ou par ajustement à qualité fixée sont proches. Elles donnent notamment les mêmes indices de Fisher. L'avantage de l'indice de Laspeyres par ajustement à qualité fixée est de ne se baser que sur une estimation de la fonction hédonique à la période de référence 0. En France, c'est le choix retenu pour l'indice des logements anciens (Gouriéroux et Laferrère 2009). La précision de ces indices est complexe à calculer. Diewert et al. (2008) note que la statistique $\left(\widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot \overline{X}^{(i)}$ peut être obtenue en régressant les valeurs $\left(\widehat{\beta}^{(t)} - \widehat{\beta}^{(0)} \right) \cdot X^{(i)}$ sur une constante pour l'ensemble des individus i de la période considérée. La variance des indices par la méthode de bien de référence se déduisent alors par Delta-Méthode. La même tech-

nique peut être adaptée aux méthodes par imputation ou à qualité fixée. Néanmoins, cela revient à considérer dans ce cas les paramètres $\widehat{\beta}^{(i)}$ comme des constantes, et donc à minorer la variance totale.

2.3.2 Le choix des caractéristiques et l'endogénéité

De manière surprenante, peu de règles sont définies dans la littérature. La règle implicite étant de faire avec ce que l'on a et que dans tous les cas, il ne faut pas chercher à expliquer les prix implicites obtenus. La théorie économique établit qu'il ne faut pas mettre de variables décrivant l'offre et la demande. La fonction de prix hédonique étant la rencontre des enchères des producteurs et des consommateurs, elle ne dit donc rien sur l'offre et la demande. Du fait de relations complexes entre cette fonction de prix et les courbes d'offre et de demande, des signes contraires à l'intuition peuvent apparaître (i.e. que consommer plus d'une caractéristique serait dévalorisée dans le prix). C'est l'analyse de Pakes (2003) qui relie fonction de prix hédonique aux coûts marginaux supportés par les firmes. Cet argument peut être renforcé par le fait que dans un cadre de pure prévision, l'endogénéité n'est pas un problème.

Cette analyse pose néanmoins question. En pratique, si la régression hédonique est utilisée pour prévoir les prix (ou les ajuster), le statisticien veut également que cette prévision ait du sens. Des signes contraires à l'intuition peuvent en effet provenir d'autres sources que celles issues de la théorie économique. La partie suivante montre le danger d'omettre des caractéristiques importantes. Le fait d'approcher certaines caractéristiques par des variables annexes peut également poser problème si ces variables peuvent être expliquées de manière simultanée au prix ou approchent également la demande ou l'offre du bien (par exemple la taille de la population d'une zone géographique). Nous privilégions l'approche prudente consistant à vérifier le signe des paramètres de la fonction hédonique.

L'approche hédonique suppose que les caractéristiques des biens sont connues à la fois des acheteurs et des vendeurs. Pour les aménités environnementales tels que les risques naturels ou technologiques, ce n'est pas toujours le cas. Il peut de plus y avoir des vices cachés. Des approches récentes (Pope 2008) montrent que lorsque l'information est incomplète, la fonction de prix hédonique est mal valorisée par l'acheteur.

2.3.3 Biais de variables omises

Benkard et Bajari (2005) analyse le biais de variables omises dans une régression hédonique $\text{Log}(\text{Prix}) = \beta \cdot X + \eta \cdot Z + \varepsilon$ où les variables Z sont inobservées. Trois sources d'erreur

TABLE 1 – Les différents indices hédoniques

Modèle	Indice	Avantages	Défauts
Agrégation de périodes adjacentes	Approché $\text{Exp}(\widehat{\delta}_t)$	<ul style="list-style-type: none"> - Simple - Econométrie spatiale possible - Adapté à de petits échantillons - Biais de non prise en compte du facteur $\text{Exp}\left(\frac{1}{2}V(\widehat{\delta}_t)\right)$ faible - Variance obtenue par Delta-Méthode $V(\widehat{\delta}_t) \text{Exp}[2\delta_t]$ 	<ul style="list-style-type: none"> - β constants - Indice variable à chaque actualisation (hors modèle sur 2 périodes adj.) - Pas de lien direct avec la théorie des indices
Imputation de prix	$L : \sum_{i \in C_0} \text{Exp}\left(\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot X^{(t)}\right)$ $P : \sum_{i \in C_t} \text{Exp}\left(\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot X^{(t)}\right)$ $F : \sqrt{L \cdot P}$ $GL : \text{Exp}\left(\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \overline{X^{(0)}}\right)$ $GP : \text{Exp}\left(\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \overline{X^{(t)}}\right)$ $T : \text{Exp}\left(\frac{1}{2}\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \left(\overline{X^{(0)}} + \overline{X^{(t)}}\right)\right)$	<ul style="list-style-type: none"> - Lien direct avec la théorie des indices - Indice stable, chaîné ou non - Meilleure prise en compte du biais de variable omise - Econométrie spatiale possible 	<ul style="list-style-type: none"> - Non prise en compte des interactions entre les périodes (Seemingly Unrelated Regressions) - Biais de non prise en compte du facteur $\text{Exp}\left(\left[\frac{\sigma_t^2}{\sigma_0^2} - \sigma_0^2\right]/2\right)$ potentiellement fort - Précision complexe à calculer - Nécessite de grands échantillons
Définition d'un bien de référence	$L : \text{Exp}\left(\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \overline{X^{(0)}}\right)$ $P : \text{Exp}\left(\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \overline{X^{(t)}}\right)$ $F : \text{Exp}\left(\frac{1}{2}\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \left(\overline{X^{(0)}} + \overline{X^{(t)}}\right)\right)$	<ul style="list-style-type: none"> - Indice stable, chaîné ou non - Meilleure prise en compte du biais de variable omise 	<ul style="list-style-type: none"> - Pas d'intégration possible des données spatiales - Lien avec la théorie des indices moins clair - Idem méthode d'imputation
Ajustement à qualité fixée	$L : \text{Exp}\left(\left(\frac{\text{Log}(\text{Prix}^{(t)})}{\text{Log}(\text{Prix}^{(0)})} - \beta^{(0)}\right) \cdot \overline{X^{(0)}}\right)$ $P : \text{Exp}\left(\alpha^{(t)} - \left(\frac{\text{Log}(\text{Prix}^{(0)})}{\text{Log}(\text{Prix}^{(0)})} - \beta^{(0)}\right) \cdot \overline{X^{(0)}}\right)$ $F : \text{Exp}\left(\frac{1}{2}\left(\frac{\beta^{(t)}}{\beta^{(0)}} - \beta^{(0)}\right) \cdot \left(\overline{X^{(0)}} + \overline{X^{(t)}}\right)\right)$	<ul style="list-style-type: none"> - Idem bien de référence - Modèle implémentable à partir de la seule période de référence 	<ul style="list-style-type: none"> - Idem bien de référence

Note de lecture : La classification, les avantages et défauts sont issus de Hill (2013). L pour indice de Laspeyres, P pour indice de Paasche, F pour indice de Fisher (i.e. $\sqrt{L \cdot P}$), GL pour indice géométrique de Laspeyres, GP pour indice géométrique de Paasche, T pour indice de Törnqvist (i.e. $\sqrt{GL \cdot GP}$).

sont en effet possibles, liées à l’erreur sur la mesure du prix (ε), aux caractéristiques inobservées ($\xi \cdot Z$) ou au choix de la forme fonctionnelle. Même dans le cas d’une variable inobservée exogène, i.e. $\mathbb{E}(Z/X) = 0$ et où la fonction de prix hédonique est non biaisée à chaque date, les indices hédoniques peuvent être biaisés pour deux raisons. Le contenu moyen des caractéristiques à chaque période t $\overline{Z^{(t)}}$ peut ne pas être le même, en cas de progrès technique non pris en compte par exemple. Le rendement des caractéristiques $\eta^{(t)}$ peut également différer selon les périodes. Benkard et Bajari (2005) proposent une analyse à facteurs pour réduire ce biais. Cette méthode nécessite néanmoins d’observer le même bien à plusieurs périodes, ce qui ne sera pas le cas dans notre application pratique.

2.3.4 La prise en compte de l’espace : autocorrélation et hétérogénéité

Pour prendre en compte l’espace (ou l’hétérogénéité des produits), la technique la plus classique consiste à segmenter le marché global, et à ajouter des indicatrices de classes (géographiques ou marques par exemple). Un indice hédonique élémentaire est alors défini par sous-marché, qui peut ensuite être agrégé. C’est le choix effectué pour les indices de logements (Bhattacharjee *et al.* 2014 pour une analyse récente des méthodes de choix de ces sous-marchés). Malgré le fait que des tests d’autocorrélation spatiale des prix de l’immobilier concluent toujours à la présence de dépendance spatiale des données (et que la littérature économique est vaste), peu d’applications de construction d’indices spatiaux existent (Dorsey *et al.* 2010 pour un exemple). Ces pistes demeurent pour l’heure encore exploratoires.

2.3.5 L’ajustement qualité

Pakes (2003) montre que les indices par appariement (“matched model”), pour lesquels les évolutions de prix sont calculés à partir des seuls biens présents à chaque période, sont biaisés. Il y a en effet un biais de sélection, les biens qui disparaissent sont ceux avec des profits à la baisse. Il est donc nécessaire dans un panier de biens de référence de substituer les biens qui disparaissent par d’autres biens. Néanmoins, ces biens de substitution seront proches mais pas exactement les mêmes. Un consommateur adaptera donc son panier de consommation pour conserver la même utilité. Consommer une unité du nouveau bien sera équivalent à consommer α unité de l’ancien bien. Si $\alpha > 1$ (resp. $\alpha < 1$), le nouveau bien sera de meilleure qualité (resp. de moindre qualité) que l’ancien. La théorie économique (Triplet 2004) nous dit alors qu’à l’équilibre un consommateur aura la même utilité à consommer une unité de l’ancien bien au prix p_1/α ($= p_0$) ou une unité du nouveau bien au prix p_1 . Toute l’information sur la qualité des biens est ainsi comprise dans les prix observés sur le marché, ce qui rejoint

les hypothèses générales sur la théorie des prix hédoniques. En particulier, si les produits sont observés simultanément, $\alpha = p_1/p_0$. En pratique, les biens sont substitués et ne sont donc pas observés simultanément à la période courante t . Plusieurs méthodes d’ajustement qualité existent. En l’absence d’informations, les plus simples consistent à considérer que la contribution à l’indice du nouveau bien est nul ($\tilde{p}_{1,t-1} = \alpha \cdot p_{0,t-1} = p_{1,t}$, avec $\tilde{p}_{1,t-1}$ le prix équivalent du bien 1 à la période $t-1$ et $\alpha = p_{1,t}/p_{0,t-1}$), ce qui revient à construire un indice par appariement, ou égale à l’évolution moyenne de l’indice pour les biens non substitués. Ces choix ne corrigent que très partiellement le biais de sélection, associé à la disparition de produits. Ils évitent que le panier de biens de références soit vide après de nombreuses périodes. C’est pourquoi lorsque les données s’y prêtent, il est préférable de mettre en oeuvre des méthodes de chevauchement (“overlapping”) ou hédoniques. Lorsque les prix de l’ancien et du nouveau bien sont observés simultanément lors de périodes précédentes ($t-1$ ou $t-2$), le coefficient de qualité α est alors le ratio des prix, $p_{1,t-1}/p_{0,t-1}$ ou $p_{1,t-2}/p_{0,t-2}$. Les prix équivalents du bien 1, $\tilde{p}_{1,t-1}$, sont alors $p_{1,t-1}$ ou $p_{1,t-2} \cdot p_{0,t-1}/p_{0,t-2}$. Sous réserve de disposer de caractéristiques observables des biens, la méthode des prix hédoniques améliore la méthode de chevauchement en supprimant l’erreur stochastique incluse dans le prix. On a alors $\alpha = \hat{p}_{1,t-1}/\hat{p}_{0,t-1}$ et $\tilde{p}_{1,t-1} = p_{0,t-1} \cdot \hat{p}_{1,t-1}/\hat{p}_{0,t-1}$. A l’aide des données de caisse, Léonard et al. 2013 montrent qu’un ajustement qualité est nécessaire pour les yoghourts et les tablettes de chocolat pour l’indice des prix à la consommation de l’INSEE. On constate que la méthode des prix hédoniques est la meilleure méthode, de faibles différences étant constatées avec les autres méthodes.

Dans le cas d’un ajustement qualité, la méthode des prix hédoniques est un outil mais ne participe pas à la construction complète de l’indice. Si les données se prêtent à estimer un modèle hédonique, un indice hédonique peut être construit. Ce sont plus des questions d’ordre pratique (nombre de caractéristiques disponibles, ou qualité d’ajustement du modèle notamment lié à l’effet magasin inobservé) qui amène à limiter la méthode à l’ajustement qualité plus qu’à la construction complète de l’indice. Les points méthodologiques précédents restent donc valables pour l’ajustement qualité.

La méthode des prix hédoniques est attractive et reste la meilleure méthode pour des biens composites. Elle suppose néanmoins de disposer de suffisamment de caractéristiques. Il faut garder à l’esprit qu’elle n’est qu’une approximation. Certaines hypothèses sont en effet peu réalistes, notamment le caractère séparable de la fonction de prix en fonction des caractéristiques, l’atteignabilité de toutes les composantes, ou l’équilibre du marché à toutes dates. Le cadre d’utilité constante est ainsi restrictif. Un approfondissement des fonctions

d'utilité des consommateurs et de production des firmes sous-jacentes aux indices construits serait nécessaire (dans l'esprit de Sillard 2013).

Nous allons illustrer les points mentionnés dans cette partie en estimant la régression de première étape des prix des terrains à bâtir. Le choix des variables explicatives, et la stabilité géographique et temporelle des résultats seront étudiés, ainsi que l'interprétation économique de cette régression clarifiée. Des indices de prix seront ensuite construits, montrant les effets d'un biais de variable omise et du type d'indice choisi.

3 Illustration à partir du prix des terrains à bâtir

3.1 Données et statistiques descriptives

Nous disposons des observations de l'enquête EPTB (Enquête sur le Prix des Terrains à Bâtir) pour les exercices 2006 à 2012, soit 662.060 données individuelles. L'enquête EPTB interroge annuellement les individus ayant obtenu un permis de construire sur les caractéristiques du terrain (prix, surface, viabilisation), le coût prévisionnel de la maison et son mode de chauffage, ainsi que les caractéristiques des individus (CSP, âge). Le questionnaire fait une page. Le champ de l'enquête se limite aux maisons neuves individuelles, ce qui exclut les rénovations et les immeubles en copropriété. L'enquête est exhaustive depuis 2010, avec un taux de réponse de 70%. Les données sont corrigées de la non-réponse partielle par imputation, pour environ un tiers des observations.

Nous excluons de notre analyse les terrains situés en Corse (moins de 1% des observations) ou dans les DOM (moins de 3%). Nous ne prenons pas en compte les terrains obtenus par donation ou héritage, faute de prix fiables (entre 14% et 21%). Les terrains achetés les années précédant le permis de construire (par exemple un terrain acheté en 2005 et un permis de construire déposé en 2006) sont également exclus de l'analyse. Ces données de prix ne sont en effet pas diffusées. Les analyses pourraient ainsi souffrir d'un biais de sélection, si par exemple les acheteurs qui décident de retarder leur construction ont un profil spécifique ou que les terrains achetés sont de nature différente. Entre 17% et 24% des observations (selon les années) sont en effet exclues par ce filtre, dont 73% pour des observations antérieures d'une année seulement au permis de construire. Enfin, les observations incohérentes (i.e. que la différence entre la surface achetée et celle déclarée sur le permis de construire est supérieure à 5%) ou atypique (i.e. jugée aberrante à partir d'une régression robuste du log du prix en fonction du log de la surface) sont exclues (moins de 5%). Notre échantillon comprend in fine

368.552 observations, avec un maximum de 71.900 observations pour l'enquête 2012 et un minimum de 28.019 observations pour l'enquête 2009 (qui était un sondage).

Des poids sont calculés pour tenir compte du plan de sondage (avant l'enquête 2011) et du calage sur marges. Les analyses descriptives sont donc pondérées. Sur la période d'étude, le nombre de permis de construire était maximum en 2006. Il a atteint un minimum en 2009 avec un indice de 59 en base 100 2006. Le nombre de transactions a ensuite augmenté après 2009 mais sans atteindre les valeurs de 2006. 96% des terrains sont utilisés pour construire une résidence principale. 63% ont été viabilisés. 86% des chefs de ménage ont moins de 50 ans, dont 27% moins de 30 ans, 41% entre 30 et 39 ans et 17% entre 40 et 49 ans. Près de 30% sont employés ou ouvriers, 13% occupent une profession intermédiaire et 17% sont cadres. 43% ne sont passés par aucun intermédiaire (agence, constructeur...) pour l'achat du terrain.

Les prix et les surfaces achetées sont décrits dans le tableau 2. Les prix ne sont pas déflatés. Le prix moyen d'un terrain est de 64.600 euros, la surface moyenne de 1144 m². Le prix au m² moyen est de 82 euros. La médiane est plus de 20% plus faible, ce qui met en avant la présence de biens très chers ou à la surface très importante. La relation entre la surface et le prix de vente est décroissante. Les biens les plus chers ne sont pas les plus grands. Cela peut s'expliquer par le fait que les plans locaux d'urbanisme définissent des surfaces maximales de construction, à travers le COS² (Coefficient d'Occupation des Sols, qui définit le nombre de m² qui peut être construit y compris les étages) et le CES (Coefficient d'Emprise au Sol, qui définit la part du terrain qui peut être construite). Le Plan Local d'Urbanisme (PLU) est défini au niveau communal ou inter-communal et inclut de multiples autres contraintes réglementaires. Il n'y a pas de base nationale qui décrit le contenu de ces PLU. On peut néanmoins penser que ces contraintes sont des facteurs importants de l'explication du prix. Nous utiliserons donc une variable approchée du COS/CES, à travers la valeur moyenne du nombre de m² construits sur la surface du terrain pour l'ensemble des terrains dans un rayon de 10 km. La surface au sol de la maison ou le nombre d'étages ne sont en effet pas connus.

L'identifiant du permis de construire permet de connaître la commune qui a délivrée ce permis. Les observations sont donc appariées avec des indicateurs d'accès aux équipements (Base Permanente des Equipements 2012), d'accès au haut-débit, ou du caractère littoral ou de haute-montagne de la commune (Observatoire des territoires), des risques naturels et industriels (GEOIDD) ou d'utilisation des terres (Corinne Land Cover). Plutôt que la

2. Le COS a été supprimé par la loi ALUR 2014.

simple présence d'équipements ou d'aménités dans la commune, des distances entre centre des communes (mesurant l'accessibilité) sont calculées.

3.2 Les déterminants des prix des terrains à bâtir

On considère ici un terrain à bâtir dont le prix s'explique par quatre types de variables :

- Les caractéristiques du terrain, la surface, son caractère viabilisé (raccordement à l'électricité, le gaz, l'eau, le téléphone et l'assainissement) ;
- Les caractéristiques de la localisation (rente foncière déduite de l'économie urbaine), la distance à la ville de 50.000 habitants la plus proche, et la tranche d'unité urbaine de la commune ;
- Les contraintes réglementaires de construction du PLU, i.e. une mesure approchée du COS/CES ;
- Les aménités environnementales et autres variables de localisation : le caractère littoral ou de haute-montagne de la commune, le taux d'équipements touristiques, la présence de haut-débit, la distance aux équipements (gare, aéroport, hôpital, supermarchés), les risques, et l'utilisation des sols ;

Concernant les aménités, les risques et l'utilisation des sols n'ont pas été retenus dans les modèles. Leur pouvoir prédictif complémentaire est faible. Les effets obtenus sont contre-intuitifs. C'est par exemple le cas des effets estimés du classement de la commune du terrain en zone à risque inondation, sismique ou incendie, qui serait corrélé positivement au prix. Cela peut tenir à des effets de composition très localisés qui compenseraient les effets négatifs afférant au risque. Ainsi, les zones concernées sont souvent caractérisées par un patrimoine paysager exceptionnel et une rareté certaine du foncier, renchérissant son prix en dépit du risque. Ces intuitions devraient faire l'objet d'une investigation ultérieure à l'aide de méthodes économétriques plus poussées ou de l'ajout de variables de contrôle mieux à même de prendre en compte ces spécificités locales. De même, certaines externalités négatives liées par exemple à la proximité immédiate d'un aéroport ne peuvent être mesurées à l'aide d'un simple indicateur de distance entre centre des communes. Enfin, ces aménités et équipements sont mesurés avec erreur car ils sont mesurés à une date fixe (2011-2012). Ainsi la présence de haut-débit est vraisemblablement sur-évaluée car mesurée en fin de période d'étude.

L'analyse du rôle des aménités environnementales et la construction des indices sont étudiés à travers la fonction de prix hédonique sous forme logarithmique. Le prix n'est pas calculé par m^2 , en l'absence de consensus sur ce point. Pour les terres agricoles, Choumert et Philénas (2014) souligne que la prise en compte du prix par hectare réduit le risque

d'hétéroscédaticité. Cavailhès et al. (2011) souligne la possibilité de biais associé à ce choix. De plus, quelle que soit la surface du terrain acheté, le nombre de m² constructible est réglementé par le COS et le CES. La notion de prix au m² n'a donc pas de sens économique claire³.

3.2.1 Le rôle des aménités environnementales

Le rôle des aménités environnementales est analysé à l'aide d'un modèle global auquel sont ajoutées des indicatrices annuelles, qui serviront à définir un indice dans la partie suivante.

Le tableau 3 décrit les effets de l'ajout de différentes variables explicatives. La première colonne n'intègre que les variables issues de l'enquête EPTB décrivant le terrain (surface et viabilisation). Ce modèle n'a qu'un pouvoir prédictif très faible, avec une R² de l'ordre de 2.5%. La colonne 2 intègre les variables de localisation, la colonne 3 la mesure approchée du COS. Le R² augmente fortement, respectivement de 28.4% et 35.8%. La dernière colonne intègre les aménités et autres variables de localisation. Malgré le fait que l'ensemble de ces variables ait un effet valorisé dans la fonction de prix, le gain sur le pouvoir prédictif est faible avec un R² de 38.8%. Les indicatrices de tranche d'unité urbaine captent en effet une part importante de ces aménités. Selon les contrôles, les effets de la surface ne sont pas les mêmes. Il apparaît qu'inclure une approximation du COS est essentiel, l'effet étant triplé avec ou sans cette variable. L'effet d'une viabilisation préalable du terrain est de l'ordre de 5%. La localisation a un effet déterminant sur le prix des terrains à bâtir. On estime que, toutes choses égales par ailleurs, une augmentation de 10 km de la distance à la ville la plus proche comptant au moins 50000 habitants (en 2010) se traduit par une baisse de l'ordre de 5% du prix du terrain. De même, le fait qu'un terrain se situe dans un grand pôle plutôt que dans la couronne d'un grand pôle se traduit, toutes choses égales par ailleurs, par une hausse d'environ 5,8 % de son prix si on tient compte des aménités et de 13.8% sans ces aménités. L'accessibilité aux grandes infrastructures de transports, à un hôpital ou à un supermarché est également déterminante. Ainsi, bien que l'effet soit d'un ordre de grandeur inférieur à 1 % par km, on estime tout de même que, toutes choses égales par ailleurs, l'éloignement de l'aéroport ou de la gare ferroviaire la plus proche se répercute à la baisse sur le prix du terrain. La distance au supermarché le plus proche est la plus valorisée avec une baisse de l'ordre de 1.5% par km. Les aménités touristiques et la présence de haut-débit sont également prises en compte dans le prix.

3. De ce fait, il n'y a pas non plus de raison de surpondérer les terrains les plus grands, qui ne sont pas forcément les plus chers. La question de représentativité par rapport à un stock de terrains ne se pose pas, seuls des flux de nouveaux biens sont ici mesurés.

Il est pertinent d'étudier la déclinaison de l'impact des différentes aménités dans le temps ou dans l'espace. En effet, l'estimation de modèles hédoniques renvoie à l'existence d'un marché unique et cohérent sur lequel porte l'estimation. Or, il est probable que les conditions d'achat et de vente sur les marchés fonciers résidentiels aient fortement changé entre 2006 et 2012 ou que, sur la période, elles varient fortement dans l'espace de la France métropolitaine. Cette analyse est conduite dans les tableaux 4 et 5. Le tableau 4 présente les résultats séparément sur les années 2006 à 2012. On observe que l'effet de la présence du haut débit dans la commune du terrain n'est plus significatif les dernières années, reflétant la diffusion progressive de cette aménité sur le territoire entre ces deux années. De manière plus générale, on constate que les effets ne sont pas stables dans le temps. Le tableau 5 présente des résultats issus d'estimations séparées sur les différentes tranches d'unités urbaines. Certaines aménités n'ont pas le même effet sur ces sous-ensembles géographiques distincts. Le caractère viabilisé d'un terrain est par exemple nul dans les grands pôles, mais de 12.4% dans les petits pôles. Certains effets sont par ailleurs contre-intuitifs, certaines caractéristiques pouvant être rares dans certaines tranches d'unités urbaines. Ces observations plaident pour une étude plus approfondie des facteurs déterminants les prix des terrains à une échelle très locale et de la façon dont ils se déclinent selon les caractéristiques des territoires.

3.2.2 Construction d'un indice de prix

Nous nous appuyons sur la première étape pour construire différents indices de prix. L'objectif est d'illustrer le biais de variable omise, et du choix de différentes méthodes d'indice hédonique. Le graphique 1 met ainsi en avant que, s'il est indispensable d'inclure des variables de localisation et réglementaire (COS), l'ajout de l'ensemble des aménités modifie peu l'indice. Cette analyse est cohérente avec le fait que les indicatrices de tranche d'unité urbaine captent une partie de ces aménités. Le graphique 2 compare pour la méthode du bien de référence les indices de Laspeyres, Paasche et Fisher, chaînés ou non. Dans la version non chaînée, l'indice de Laspeyres représente une borne haute et l'indice de Paasche une borne basse. Pour les indices chaînés, les 3 indices sont confondus. Cette constatation amènerait à privilégier les indices chaînés, plus stables et en cohérence avec les analyses par année de la partie précédente. L'hypothèse des indices non chaînés d'un unique ajustement des quantités consommées par les prix nous paraît de plus trop restrictive. Le graphique 3 compare les indices chaînés et non chaînés par méthode, ainsi qu'une version chaînée corrigée du facteur $\text{Exp} \left(\left[\widehat{\sigma}_t^2 - \widehat{\sigma}_{t-1}^2 \right] / 2 \right)$. Les indices chaînés sont proches des indices non chaînés, tout en les sur-évaluant de manière systématique. La correction apportée liée aux variances résiduelles

est forte. Cette correction n'est pas nécessaire pour l'indice sur période adjacente. Or, dans le graphique 4, on constate que les méthodes non corrigées de bien de référence (ou à qualité fixée) et sur périodes adjacentes donnent des résultats presque identiques. Prendre en compte un facteur correctif, établi sous une hypothèse elle-même contestable de normalité et d'homoscédasticité des termes d'erreur, apparaît ainsi comme un remède pire que le mal. Enfin, on constate une divergence entre la méthode par imputation des autres méthodes, ce qui pourrait s'expliquer par le caractère arithmétique de l'indice par imputation. Le graphique 5 compare les taux de croissance des différentes méthodes. Une fois contrôlée par les différentes variables explicatives, l'indice de prix des terrains semble refléter des hausses plus élevées en début de période (un effet qualité négatif) mais moins élevés par la suite. Avant la crise, cela peut tenir au fait que, marginalement, les derniers terrains mis en vente sont de qualité décroissante, l'inclusion de variables propres à la localisation du terrain permettant de prendre en compte cet effet en raisonnant "à qualité constante". Après la crise, cela peut traduire un effet volume, seuls les terrains de meilleur qualité trouvant acheteurs. On constate une forte divergence de la méthode d'imputation en 2009 par rapport aux autres méthodes.

4 Conclusion et perspectives

Cette étude a permis de dégager quelques bonnes pratiques sur la construction d'indices hédoniques, sur le choix des variables et des méthodes. Elle pourra être approfondie par une réflexion sur la spécification, en utilisant des approches Box-Cox, et semi-paramétrique par splines. Pour prendre en compte l'hétérogénéité géographique des résultats, nous utilisons des indicatrices de tranche d'unité urbaine dans un modèle national. D'autres pistes peuvent être envisagées, d'indices désagrégés à la régression géographique pondérée.

5 Bibliographie

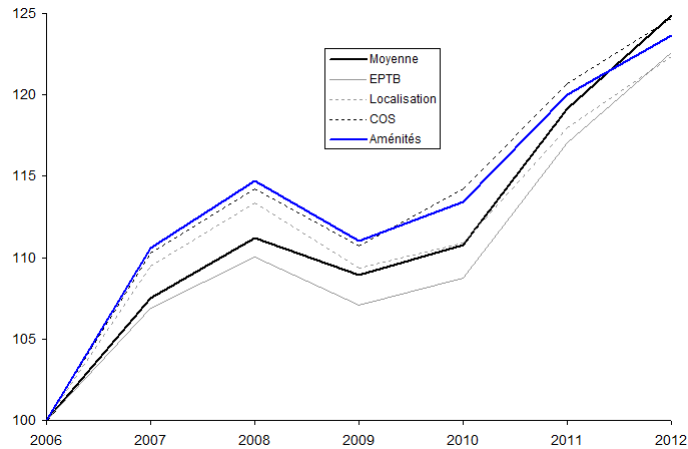
- Balcone**, Thomas. (2013) "La Construction d'un Indice des Prix des Logements Neufs" Mimeo.
- Baudry** Marc, Alain Guengant, Sophie Larribeau, et Matthieu Leprince. (2009) "Formation des Prix Immobiliers et Consentements à Payer pour une Amélioration de l'Environnement Urbain : l'Exemple Rennais." *Revue d'Economie Régionale & Urbaine*, 2, 369-411.
- Benkard**, C. Lanier, et Patrick Bajari. (2005) "Hedonic Price Indexes With Unobserved Product Characteristics, and Application to Personal Computers." *Journal of Business & Economic Statistics*, 23(1), 61-75.
- Bhattacharjee**, Arnab, Eduardo Castro, Taps Maiti, et João Marques. (2014) "Endogenous spatial structure and delineation of submarkets : A new framework with application to housing markets" Discussion Paper Series N°5, The Spatial Economics and Econometrics Centre, Heriot Watt University.
- Bureau**, Benjamin, et Matthieu Glachant. (2010) "Evaluation de l'impact des politiques "Quartiers Tranquilles" et "Quartiers Verts" sur les prix de l'immobilier à Paris." *Economie et Prévision*, 192, 27-44.
- Cavailhès**, Jean. (2005) "Le Prix des Attributs du Logement." *Economie et statistique*, 381-382, 91-123.
- Cavailhès**, Jean, Mohamed Hilal, et Pierre Wavresky. (2012) "L'Influence Urbaine sur le Prix des Terres Agricoles et ses Conséquences Pour l'Agriculture." *Economie et Statistique*, 444-445, 99-125.
- Choumert**, Johanna, et Muriel Travers. (2010) "La Capitalisation Immobilière des Espaces Verts dans la Ville d'Angers : Une Approche Hédoniste." *Revue économique*, 61(5), 821-836.
- Choumert**, Johanna, et Pascale Phélinas. (2014) "Determinants of agricultural land values in Argentina." Journées de Microéconomie appliquée (Clermont-Ferrand), Communication.
- Dachary-Bernard**, Jeanne, Frédéric Gaschet, Sandrine Lyser, Guillaume Pouyanne, et Stéphane Virol. (2011) "L'Impact de la Littoralisation sur les Marchés Fonciers Une Approche Comparative des Côtes Basque et Charentaise." *Economie et Statistique*, 444-445, 127-154.

- Diewert**, Erwin, Saeed Heravi, et Mick Silver. (2008) "Hedonic Imputation Versus Time Dummy Hedonic Indexes." NBER Working Paper N°14018.
- Dorsey**, Robert Earl, Haixin Hu, Walter James Mayer, et Hui-chen Wang. (2010) "Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle" *Journal of Housing Economics*, 19(2), 87-105.
- Fack**, Gabrielle, et Julien Grenet. (2010) "When do better schools raise housing prices? Evidence from Paris public and private schools." *Journal of Public Economics*, 94, 59-77.
- Gouriéroux**, Christian, et Anne Laferrère. (2009) "Managing Hedonic Housing Price Indexes : the French Experience." *Journal of Housing Economics*, 18(3), 206-213.
- Grislain-Letrémy**, Céline, et Arthur Katosky. (2013) "Les Risques Industriels et le Prix des Logements." *Economie et Statistique*, 460-461, 79-106.
- Hill**, Robert J. (2013) "Hedonic Price Indexes For Residential Housing : A Survey, Evaluation And Taxonomy." *Journal of Economic Surveys*, 27(5), 879-914.
- Lancaster**, Kelvin J. (1966) "A New Approach to Consumer Theory." *The Journal of Political Economy*, 74(2), 132-157.
- Lefebvre**, Lise, et Céline Rouquette. (2012) "Les Prix du Foncier Agricole sous la Pression de l'Urbanisation." *Economie et Statistique*, 444-445, 155-180.
- Léonard**, Isabelle, Patrick Sillard, et Gaëtan Varlet. (2013) "Scanner data and quality adjustment - The French Experience" Mimeo.
- Moulton**, Brent R. (2001) "Le Rôle Croissant des Méthodes Hédonistes dans l'Établissement des Statistiques Officielles aux Etats-Unis." OCDE STD/NA(2001)21, Réunion d'Experts en Comptabilité Nationale.
- Pakes**, Ariel. (2003) "A Reconsideration of Hedonic Price Indexes with an Application to PC's." *American Economic Review*, 93(5), 1578-1596.
- Palmquist** R. B. (2005) "Property Value Models." Chapter 16 in *Handbook of Environmental Economics*, Vol. 2, pp.764-819, Edited by K.-G. Mäler et J.R. Vincent.
- Pope**, Jaren C. (2008) "Do Seller Disclosures Affect Property Values? Buyer Information and the Hedonic Model." *Land Economics*, 84(4), 551-572.
- Rosen** S. (1974) "Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition." *The Journal of Political Economy*, 82(1), 34-55.

- Sheppard**, Stephen. (1999) “Hedonic Analysis of Housing Markets” Chapter 41 in *Handbook of Regional and Urban Economics*, Vol. 3, pp. 1595-1635, Edited by P. C. Cheshire & E. S. Mills.
- Sillard**, Patrick. (2013) “Les Données de Caisse : Vers des Indices de Prix à la Consommation à Utilité Constante.” Document de travail INSEE N°F1305.
- Triplett**, Jack. (2004), “ Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes : Special Application to Information Technology Products. ” OECD Science, Technology and Industry Working Papers, 2004/09, OECD Publishing.

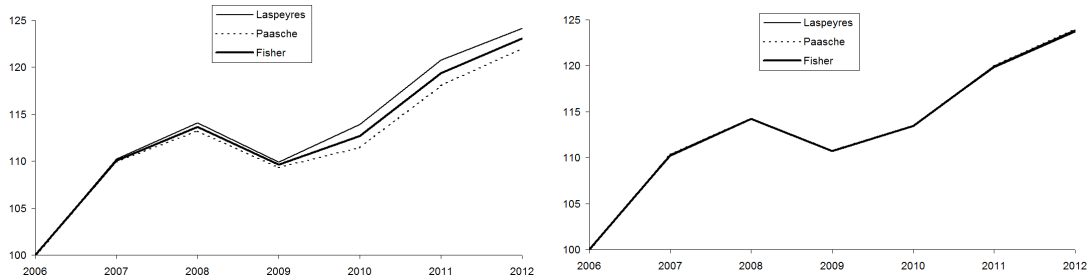
6 Tableaux et graphiques

Graphique 1 : Indices de prix, selon les variables explicatives prises en compte



Note de lecture : L'indice non corrigé de la qualité est représenté par la ligne noire pleine. Les autres indices sont corrigés de la qualité, à l'aide d'indicatrices d'année dans une régression hédonique. L'indice 'EPTB' (ligne grise pleine) inclut la viabilisation et la surface comme variables d'ajustement qualité, l'indice 'Localisation' (ligne grise pointillée) la distance à la commune de 50.000 habitants la plus proche et des indicatrices de tranche d'unité urbaine, l'indice 'COS' (ligne noire pointillée) une valeur approchée du coefficient d'occupation des sols et l'indice 'Aménités' (ligne bleue pleine) l'ensemble des aménités et autres variables de localisation.

Graphique 2 : Indices de Laspeyres, Paasche et Fisher pour un bien de référence

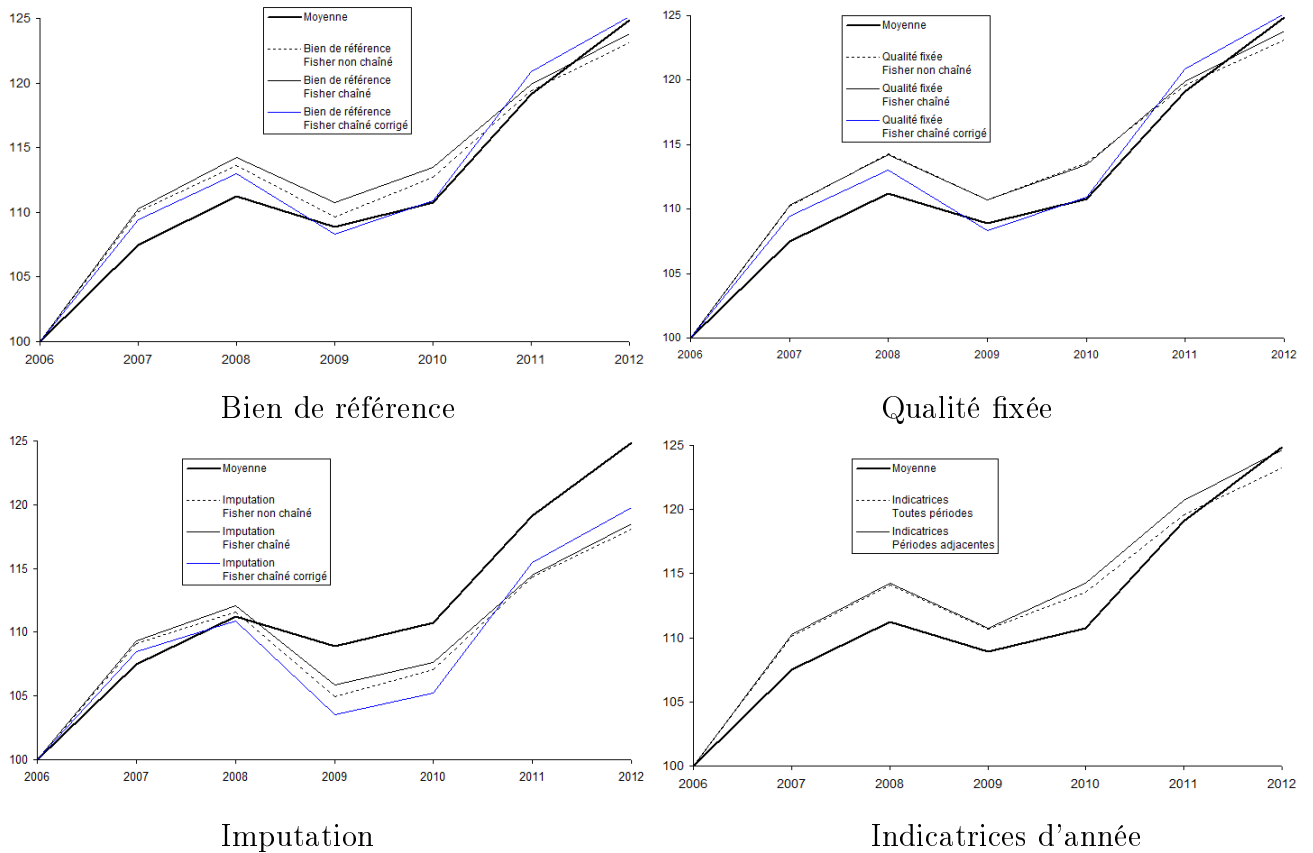


Indice Non Chaîné

Indice Chaîné

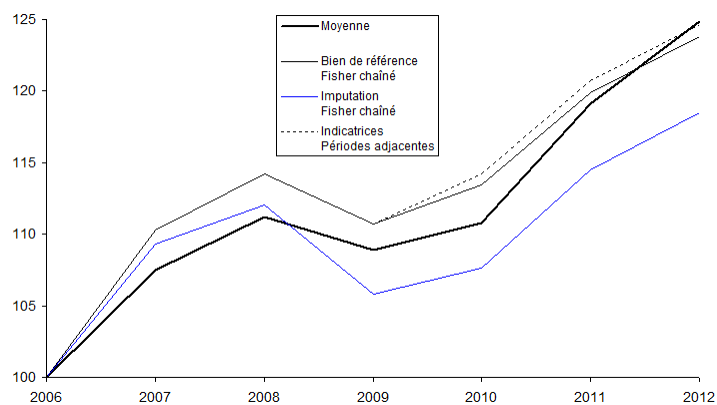
Note de lecture : Les indices sont calculés à l'aide de la méthode dite de bien de référence, défini par \bar{X}_{2006} pour l'indice de Laspeyres, \bar{X}_{2012} pour l'indice de Paasche. Les variables explicatives incluent les variables issues de l'enquête EPTB, de localisation et une approximation du COS.

Graphique 3 : Indices de prix chaînés ou non, selon les méthodes



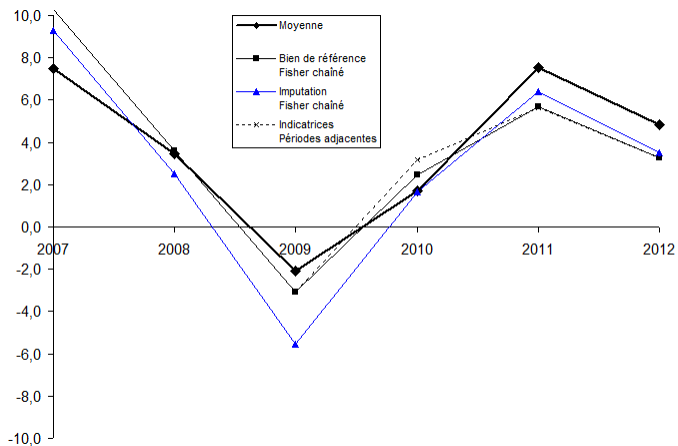
Note de lecture : L'indice non corrigé de la qualité est représenté par la ligne noire pleine. Les indices chaînés sont représentés par une ligne noire pleine fine, les indices non chaînés par une ligne noire pointillée. Une version corrigée à l'aide du facteur $\text{Exp}\left(\frac{[\widehat{\sigma}_t^2 - \widehat{\sigma}_{t-1}^2]}{2}\right)$ des indices chaînés de Fisher est représentée par une ligne pleine bleue. Les variables explicatives incluent les variables issues de l'enquête EPTB, de localisation et une approximation du COS.

Graphique 4 : Comparaison des indices de prix chaînés selon les méthodes



Note de lecture : L'indice non corrigé de la qualité est représenté par la ligne noire pleine. Les autres indices sont corrigés de la qualité, à l'aide de différentes méthodes. La méthode à qualité fixée n'est pas représentée car les résultats sont identiques à ceux de l'indice obtenu avec la méthode du bien de référence. Les variables explicatives incluent les variables issues de l'enquête EPTB, de localisation et une approximation du COS.

Graphique 5 : Comparaison des taux de croissance des indices de prix chaînés selon les méthodes



Note de lecture : L'indice non corrigé de la qualité est représenté par la ligne noire pleine. Les autres indices sont corrigés de la qualité, à l'aide de différentes méthodes. La méthode à qualité fixée n'est pas représentée car les résultats sont identiques à ceux de l'indice obtenu avec la méthode du bien de référence. Les variables explicatives incluent les variables issues de l'enquête EPTB, de localisation et une approximation du COS.

Tableau 2 : Caractéristiques des données EPTB

	N	Moyenne	Médiane	Écart-type	Min.	Max.
Prix du terrain	368.552	64600	52731	55794	4106	487090
Superficie du terrain	368.552	1144	884	1121	107	11423
Prix au m2	368.552	82	61	100	1	2146

Note de lecture : Toutes les données du champs d'étude 2006-2012 sont utilisées.

Tableau 3 : Régression hédonique globale

	(1) Log(P) EPTB	(2) Log(P) Loc.	(3) Log(P) COS	(4) Log(P) Aménités
Surface (x10-3)	-0.068	0.041	0.115	0.119
viabilisé	0.057	0.058	0.053	0.052
Distance commune 50.000 hab.		-0.009	-0.005	-0.005
Grand pôle (GP) (Ref. Couronne GP)		0.248	0.138	0.058
Proche plusieurs GP		-0.131	-0.100	-0.089
Petit pôle (PP)		-0.167	-0.135	-0.206
Couronne PP		-0.479	-0.400	-0.408
Hors pôles		-0.406	-0.339	-0.328
COS			0.046	0.039
Haut débit				0.020
Haute montagne				0.452
Plage				0.094
Taux équipements touristique				0.038
Dist. Littoral <30 kms				-0.018
Dist. Littoral ² <30 kms (x10-3)				0.649
Dist. Gare				-0.002
Dist. Aéroport				-0.002
Dist. Hôpital				-0.003
Dist. Supermarchés				-0.014
Constante	10.810	11.037	10.258	10.487
Observations	368537	368537	368537	367159
R2 Ajusté	0.025	0.284	0.358	0.388

Note de lecture : Toutes les variables sont significatives à 1%.

Tableau 4 : Régression hédonique par année

	(1) Log(P) 2006	(2) Log(P) 2007	(3) Log(P) 2008	(4) Log(P) 2009	(5) Log(P) 2010	(6) Log(P) 2011	(7) Log(P) 2012
surface (x10-3)	0.105	0.107	0.113	0.0805	0.142	0.159	0.136
viabilisé	0.042	0.043	0.058	0.073	0.074	0.051	0.050
Distance commune 50.000 hab.	-0.006	-0.006	-0.005	-0.004	-0.0041	-0.005	-0.006
cos	0.045	0.044	0.045	0.034	0.034	0.035	0.033
Grand pôle (GP) (Ref. Couronne GP)	0.077	0.036	-0.004 (NS)	0.048	0.049	0.064	0.085
Proche plusieurs GP	-0.105	-0.109	-0.103	-0.060	-0.065	-0.083	-0.078
Petit pôle (PP)	-0.210	-0.252	-0.217	-0.185	-0.179	-0.197	-0.192
Couronne PP	-0.451	-0.415	-0.423	-0.333	-0.383	-0.396	-0.399
Hors Pôles	-0.388	-0.351	-0.332	-0.288	-0.290	-0.305	-0.285
Haut débit	0.040	0.027	0.021	0.015 (FS)	0.017	0.01 (FS)	0.009 (FS)
Haute montagne	0.521	0.499	0.452	0.444	0.354	0.407	0.433
Plage	0.100	0.145	0.077	0.056	0.078	0.064	0.094
Taux équipements touristique	0.045	0.043	0.037	0.029	0.031	0.038	0.033
Dist. Littoral <30 kms	-0.019	-0.019	-0.023	-0.016	-0.016	-0.019	-0.017
Dist. Littoral^2 <30 kms (x10-3)	0.652	0.689	0.909	0.553	0.510	0.659	0.558
Dist. Gare	-0.002	-0.003	-0.003	-0.002	0.000 (NS)	-0.001 (FS)	-0.001 (NS)
Dist. Aéroport	-0.002	-0.002	-0.002	-0.003	-0.001	-0.001	-0.001
Dist. Hôpital	-0.002	-0.004	-0.003	-0.002	-0.003	-0.003	-0.003
Dist. Supermarchés	-0.015	-0.015	-0.014	-0.014	-0.009	-0.014	-0.015
Constante	10.478	10.587	10.585	10.658	10.563	10.642	10.731
Observations	69859	59225	43193	27890	35890	59419	71683
R2 Ajusté	0.418	0.410	0.401	0.359	0.339	0.362	0.358
Variance	0.321	0.306	0.285	0.241	0.196	0.212	0.234

Note de lecture : NS pour non significatif à 10%, FS pour non significatif à 1%.

Tableau 5 : Régression hédonique par tranche d'unité urbaine

	(1) Log(P) GP	(2) Log(P) C. GP	(3) Log(P) P. GP	(4) Log(P) PP	(5) Log(P) C. PP	(6) Log(P) HP
Surface (x10-3)	0.205	0.112	0.121	0.116	0.129	0.097
Viabilisé	-0.001 (NS)	0.056	0.093	0.081	0.124	0.071
Distance commune 50.000 hab.	-0.007	-0.005	-0.007	-0.002	-0.002	-0.003
cos	0.035	0.035	0.038	0.048	0.054	0.056
Haut débit	0.008 (NS)	0.072	-0.032	0.003 (NS)	0.022 (NS)	-0.071
Haute montagne	0.158	0.403	0.418	0.329	0.283	0.498
Plage	0.014 (NS)	0.101	0.090	0.200	0.000 (NS)	0.155
Taux équipements touristique	0.049	0.033	0.031	0.065	0.046	0.025
Dist. Littoral <30 kms	-0.014	-0.031	-0.032	0.000 (NS)	0.023	0.002 (FS)
Dist. Littoral ² <30 kms (x10-3)	0.289	1.161	1.350	-0.226	-0.827	-0.070 (FS)
Dist. Gare	-0.003	-0.003	0.002	-0.002	-0.003	0.000 (NS)
Dist. Aéroport	-0.002	-0.001	-0.003	-0.003	0.000 (NS)	-0.001
Dist. Hôpital	0.007	-0.004	-0.004	-0.003	-0.002 (FS)	-0.003
Dist. Supermarchés	0.000 (NS)	-0.019	-0.014	0.005 (FS)	-0.022	-0.010
Constante	10.653	10.554	10.494	10.010	9.547	9.730
observations	79965	147355	37768	28967	6395	66709
R2 Ajusté	0.312	0.247	0.277	0.275	0.188	0.252

Note de lecture : NS pour non significatif à 10%, FS pour non significatif à 1%. GP pour Grand Pôle, C.GP pour Couronne de Grand Pôle, P.GP pour Proche de plusieurs Grands Pôles, PP pour Petit et moyen Pôles, C.PP pour Couronne de Petit et moyen Pôles, HP pour Hors Pôles.